Contributions of academic articles to the practice of journalism and data management

Idoia Portilla*

* Universidad de Navarra, Spain

Abstract

The increasing number of open data resources and the arrival of big data have boosted the data available as a source of news. Journalists need new skills for collecting the data and creating the news. In addition, all this data can also be used to provide new media services and to take media business decisions, and journalists need new skills related to data for these tasks. Taking into account these areas of knowledge required by journalists for the use of data, we perform a structured literature review (SLR) followed by a content analysis. The results confirm the relevance of data management in journalistic practice, requiring skills in statistics, data visualization, technology, but also in ethics, marketing or audience monitoring.

Keywords: skills, big data, management, content analysis, visualization.

Introduction

The relevance of acquiring skills for data management in journalism practice has been boosted by the increasing number of open data resources and the arrival of big data (Elías-Pérez, 2015; Parasie, 2015). At the same time, the media capable of working with data can generate databases that can be sold or used to offer extra services (Gray, Bounegru, and Chambers, 2012). This is the case, for example, of *ProPublica* (www.propublica.org/datastore/) or *Bloomberg for enterprise* (www.bloomberg.com/enterprise/).

But data is not only a product or service for the media. Nowadays, all digital business generates a great amount of data, and the proper use of this big data can improve company results (Mayer-Schönberger and Cukier, 2013; Lewis and Westlund, 2015). It is the era of big data: high-volume of data in high-variety information assets, that is generated at high-velocity and requires being analysed also very quickly to gain valuable insights. As other media companies with big data, media companies should learn how to monetize the data generated by the digitalization of their content, taking advantage of information such as the visits to their webs, the fans of their brand or the searches made in their cross-platform content services.

In response to these realities, journalists should gain new competencies related to the management of data. In this paper, we review the milestones linking data with journalism practice. The first step is obtaining the main areas of knowledge required by journalists in order to use data. The second step is performing a quantitative literature review of research contributions published in the main academic journals. The results of this review, complemented by a content analysis, will provide a list of areas of knowledge and skills that are becoming important requirements for the journalist in this era of big data.

Milestones related to data and journalism

Journalists have been using data in economic or sports news since the origins of media. The first periodical publication in the eighteenth century, *The Lloyd's List*, was a weekly journal of general commercial news and included price lists and details of ships arriving at English and Irish ports (Arrese, 2002). According to *The Guardian*, their first data journalism content was published in 1821 with a list of schools in Manchester and Salford, how many pupils attended each one and their average annual spending (Datablog, 2011).

Data continued being source and content of news in the following decades, but the arrival of computers provided a great step forward. In 1952, the US television network CBS used a computer to predict election results (Hewett, 2013; Uskali and Kuutti, 2015). In 1967, the coverage of the 1967 Detroit race riots by Phillip Meyer won the Pulitzer Prize (Meyer, 2001; Uskali and Kuutti, 2015). This last journalist is considered the first American journalist to have used computers to conduct investigative journalism (Parasie and Dagiral, 2013), and many authors reference his work in the history of data journalism (Hughes, 2013; Bradshaw, 2014; Elías-Pérez, 2015).

However, many journalists were not capable of analysing and reporting data, so they needed training. Meyer himself was the author of the handbook *Precision journalism* published in the 70's (Meyer, 2001; Coddington, 2015). In his book, he tried to show media professionals how to deal with data and surveys in their reports (Meyer, 2001). In the following years, we can find other handbooks about computer reporting (Houston, 1996; DeFleur, 1997; Garrison, 1998; Houston, Bruzzese, and Weinberg, 2002), showing the relevance of the issue. The set-up of the National Institute for Computer-Assisted Reporting in the United States in 1989 (Hughes, 2013; Parasie and Dagiral, 2013) and the Centre for Investigative Journalism (CIT) in 2003 in the UK (Hewett, 2013) also helped train journalists in data management.

From the mid-1990s, the number of data-oriented projects increased thanks to the growing availability of computerized data and the world wide web (Parasie, 2015). In 2010, *The Guardian* in the UK and the *New York Times* and the Chicago Tribune in the United States developed their own data journalism teams and started publishing news with the data they elaborated (Uskali and Kuutti, 2015). In 2012 the Global Editors Network presented the first Data Journalism awards (Uskali and Kuutti, 2015).

The educational resources teaching how to use data as a source, to collect it, to analyse it and to report it, have been included under several terms: precision journalism, Computer Assisted Reporting (CAR) or investigative journalism (Meyer, 2001; Coddington, 2015; Uskali and Kuutti, 2015). The term "data journalism" has generally been used since the mid-2000s (Royal and Blasingame, 2015). However, it is still difficult to define (Ferreras-Rodríguez, 2012; Beleaga, 2013; Coddington, 2015; Royal and Blasingame, 2015). Professionals and scholars sometimes separate data journalism from other practices such as computational journalism, programmer journalism or open-source journalism (Coddington, 2015).

If we consider any kind of journalism that uses data as a source for its news and that publishes stories using data, we can divide its practice into three steps: access to the data, its analysis and the publication of the story (Ferreras-Rodríguez, 2012; Teruel and Blanco, 2015). If we consider the analysis as a skill necessary for reporting, two relevant areas of knowledge stand out over the rest: data as a source of news and the use of data in the content of news.

Open data and software

Other milestones related to data and journalism are the Open Systems Movement in the late 1970s (Holtgrewe and Werle, 2001), the open source principle born in the late 1990s (Coddington, 2015) and the open data policies (Hughes, 2013). These open movements have facilitated the access to data and the software for its analysis (Ferreras-Rodríguez, 2012).

The passing of the Freedom of Information Act in the United States was in 1966 (Bradshaw, 2014), and it gave people the right to access information from the government. However, the release of databases by government authorities, the actual open data policies, started in the 1980s (Parasie and Dagiral, 2013). In the UK, the Freedom of Information Act received royal assent in 2000; it came fully into force in 2005 (Davies, 2010; Hewett, 2013; Bradshaw, 2014) and, in 2010, the government open data website *data.gov.uk* was launched (Davies, 2010; Hewett, 2013), a few years after *The Guardian* journalists´ campaign called Free Our Data (Davies, 2010; Hewett, 2013; Bradshaw, 2014). In the European Union, regulation 1049/2001 is the transparency law, although other countries have already tried to regulate it: Sweden, in 1776, Finland in 1951, and Spain in 1992 (La-Rosa and Sandoval-Martín, 2016).

The Open Government Partnership was launched in 2011 (Bradshaw, 2014). At that time, governments began to become members so as to demonstrate their compromise with transparency and the empowerment of citizens (Open Government Partnership, 2016). Regulators, consumer groups, charities, scientific institutions and businesses also began releasing newsworthy data (Bradshaw, 2010), making the analysis of data among journalists more common.

But the use of government data for developing news content has been seen as the beginning of a new kind of data journalism, mainly due to Wikileaks. This release of war logs from Iran and Afghanistan in 2010 is considered a significant event for data journalism (Hewett, 2016), and even the starting point for contemporary data journalism activities (Uskali and Kuutti, 2015). In 2013, when Edward Snowden revealed that intelligence agencies were engaging in secret global surveillance, the debate about privacy moved to the foreground (McStay, 2017; Lewis and Westlund, 2015). Thus, the 2010s are relevant dates for data and journalism practice.

In brief, these open data and open source movements helped journalists in the accessing of data and its analysis for reporting a story, again bringing these two areas to the forefront, now together with the privacy debate.

Portals and services based in data

Besides data as a source, the open data movement provided portals and services based on data. In 1998 the UK's *mySociety* developed *Up My Street*, a website with information of local areas (Bradshaw, 2014). In 2004, Washington DC launched its data portal (Parasie and Dagiral, 2013). These portals facilitated access to data for news, but also the development of new media services mainly as public services for their audiences.

Since 2005, journalists have been hired to design web applications, with the purpose of strengthening government accountability and citizen participation (Parasie and Dagiral, 2013). In 2007, the Knight

Foundation awarded the project *EveryBlock*, a website with information about neighbourhoods. The project *PolitiFact*, a fact-checking website that rates the accuracy of claims by US politics, received a Pulitzer award for its journalist content based on data and analytics (Ferreras-Rodríguez, 2012).

Some data services, such as the BBC budget and fuel calculator, have helped UK citizens to control their expenses (Hurrel and Walton, 2013). Others, such as mashups and web applications, make data transparent to the audience and can be used as decision-making tools, for example, to decide where to live, knowing whether the neighbourhood is a secure area (Parasie and Dagiral, 2013). So, online data services are seen as an entrepreneur model of data journalism activities in the newsroom, according to Uskali and Kuutti (2015), with stories that incorporate user participation (Royal and Blasingame, 2015), and look to help people to become more powerful.

As we have seen in the introduction, some media, such as *ProPublica* or *Bloomberg* generate databases to be sold (Gray, Bounegru, and Chambers, 2012). There are also news organizations that employ data analytics as their niche, such as Nate Silver`s *FiveThrityEight* (fivethirtyeight.com) in the United States (Royal and Blasingame, 2015), or *Fundación Civio* (www.civio.es) and *Porcentual* (www.porcentual.es) in Spain.

Nowadays, the increasing number and use of mobile phones has forced media to offer their services via apps. News applications can personalize data to the user offering information about traffic or sports (Westlund, 2013; Howard, 2014; Silva, López, Westlund and Ulloa, 2016).

According to all these examples, apart from data as a source of news and as content of them, a third area to be considered is the use of data to offer media services, which can provide new revenue resources for media companies so journalists should be aware of it.

The era of Big Data

The development of big data is the last milestone we highlight. This term was coined in the 2000s by sciences such as astronomy and genomic, and it migrated to all areas about ten years later (Mayer-Schönberger and Cukier, 2013).

Big data can be defined as the data generated through the use of digital technologies (Margetts, 2014) and since 2001 it has been described by three Vs: volume, velocity, and variety (Laney, 2001). It can also refer to the processes surrounding this great volume of data, which is too large for standard computer memory and software (Lewis and Westlund, 2015; Parasie, 2015).

Due to the increase in digital information which led to the big data era, data-driven journalism has become more prominent (Coddington, 2015). According to Elías-Pérez (2015), the computer experts know how to manage data, but the journalists are the ones who know what questions should be asked to tell the story. Therefore, big data affects data journalism in sources used and in reporting. Now it is easier to find the data, but the journalist needs to select, analyse and compare it and finally publish the news (Ferreras-Rodríguez, 2012).

Furthermore, big data relates to business. Digitalization has given media companies computer data that includes information about the user (Manovich, 2001; Napoli, 2016) and can drive the online business of the media (Aguado and Martínez, 2009; Lewis and Westlund, 2015; Royal and Blasingame, 2015). The most

used methods of capturing data are online tracking cookies and website pathing (Forbes Insights, 2014) but also mobile apps can provide useful information (Howard, 2014). Data can be used to customize automatically the media composition, personalizing the offer (Evens and Van Damme, 2016). It offers great opportunities to the communication market (Marks, 2013; Papí-Gálvez, 2014). The media can implement big data strategies for a better understanding and targeting of customers (Napoli, 2016).

In brief, the big data era means that the management of data is essential for journalists because the data available is increasing and news stories should take this into account. However, media companies have also gained big data related to their business, so journalists should learn how to manage it for decision-making.

From the milestones to areas linking data and journalism

Taking into account the development of data journalism and the aforementioned milestones related to data and journalism practice, we will consider four areas that linked journalism and data:

1. Data as a source of news.

   Data has been the source of news since the origin of the media but journalists received real training in the collection of data with the Meyer's precision journalism book and the development of Computer Assisted Reporting (CAR). With the open data movements, access to data became easier, and nowadays, in the era of big data, journalists can find more and more data that sometimes require computational knowledge in order to retrieve them.

2. Data in the content of news (reporting).

   Once journalists have the data, they need to analyse it and write the story. The educational resources under the names of precision journalism, investigative journalism, computer-assisted reporting or data journalism have tried to teach how to do this. In addition, the open software movement has helped by facilitating access to analytics and visualization software. In the era of big data, journalists have data to complement many of their news stories.

3. Data for providing new media services.

   The open data movement gave access to government data and media have used it to develop sites, portals and apps based on this information. With these innovative services, media have tried to facilitate access to data for citizens and obtain new revenues. The big data available can help to increase the number of these services.

4. Data in the media business.

   Digitalization has provided big data for online media. Journalists and managers of the media have to use data to improve their products - by providing personalized news, for example. Like other companies with digital data, the media should try implementing big data strategies to increase their revenues.

Once we have identified the areas that link data with journalism and journalists, the next step is to analyse whether these issues and milestones appear in academic publications, when and related to which skills or other issues. As open data and big data have boosted the volume of data available, the period considered in our research will include these two milestones.

Method

The methodology applied in this work is a structured literature review (SLR) followed by a content analysis with first, a priori coding and, in a second step, an emergent coding. The result is the combination of methods used in health and the social sciences.

According to Dawidowicz (2010, 5), "a literature review is an examination of scholarly information and research-based information on a specific topic". When it uses rigid rules is called a structured literature review (SLR) (Massaro, Dumay, and Guthrie, 2016). The use of rules offers less bias, more transparency, and the possibility of replication of the review (Massaro, Dumay, and Guthrie, 2016; Tranfield, Denyer, and Smart, 2003).

The SLR is frequently applied to develop insights and future research paths or review methodologies applied (Moher et al., 2009; López-Cantos, 2015; Massaro, Dumay, and Guthrie, 2016; Ausserhofer et al., 2017). However, in this case the objective is translating areas of research into skills related to journalism practice. The process for the selection of contributions will be drawn using the flow diagram proposed by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) from the field of health (Liberati et al., 2009; Moher et al., 2009). This diagram helps in explaining the process of the identification of records, the screening, the eligibility and the final records included (Moher et al., 2009).

After the selection of cases, the data will be coded to develop insights (Massaro, Dumay, and Guthrie, 2016). This codification follows the process of a content analysis (Wimmer and Dominick, 2011), registering items for each article (which constitute the unit of analysis). The content categories will be established using a priori coding for those that can be listed before the analysis. In a second step, the categories will come from an emergent coding, establishing categories after the examination of the articles (Wimmer and Dominick, 2011).

Research questions

The final purpose of this article is to know what can be learnt about skills related to the use of data in journalism from the contributions of researches published in academic journals. There have been previous literature reviews for identifying influential publications and gaps in the research of data journalism (Ausserhofer et al., 2017), for identifying methodologies (López-Cantos, 2015) or opportunities for innovation in computational journalism (Diakopoulos, 2012). However, this paper differs from these previous contributions in the method used for the selection of the contributions and the research questions made. Firstly, it extends the interest to any article that relates data and journalism/journalists. Secondly, the

questions refer to the issues discussed by academia with the focus on the skills and knowledge needed in the practice of journalism.

The research questions are:

- Which attention do the issues of (1) data as a source of news, (2) data in the content of news, (3) data for providing new services and (4) data in the media business receive in the academic contributions?
- Are any of the milestones "open data" and "big data" cited in the academic contributions?
- What journalistic skills related to data can be derived out of these academic contributions?
- Are there other relevant issues related to data and journalism/journalists in these academic contributions?

Universe of study and final sample

The unit of analysis will be the academic contributions connected with data and journalism or journalist published from January 2006 to December 2016. This period includes the years of the development of open data and big data, the two milestones of interest. The key criterion for the selection of academic contributions is that they should include the terms "data" and "journalism" or "journalist" in their abstracts. The selection is made using the tool Unika, from the library of the Universidad de Navarra (www.unav.edu/en/web/biblioteca). This search tool has been in operation since September 2011 (www.unav.edu/web/vida-universitaria/detalle-noticia-pestana/2011/09/28/unika:-nuevo-buscador-unificado-de-la-biblioteca-de-la-universidad-de-navarra?articleId=302098). It provides access to several academic journal databases. In 2017 this tool included more than 350 providers such as Academic OneFile, BASE, Biography in Context, Business Insights (Essentials), Business Source Ultimate, Expanded Academic ASAP, General OneFile, General Reference Center Gold, InfoTrac Computer Database, InfoTrac Newsstand, InfoTrac Student Edition, NewsBank and Scopus.

The Unika search tool provides information on the contribution including the title of the article, the authors, the journal, volume, issue and pages, the keywords (from the authors or by the database) and, in many cases, the abstract. The search took place on April 25 2017. The date is relevant because the journals included in each database can vary, and also the contract between the library and the providers can change over time.
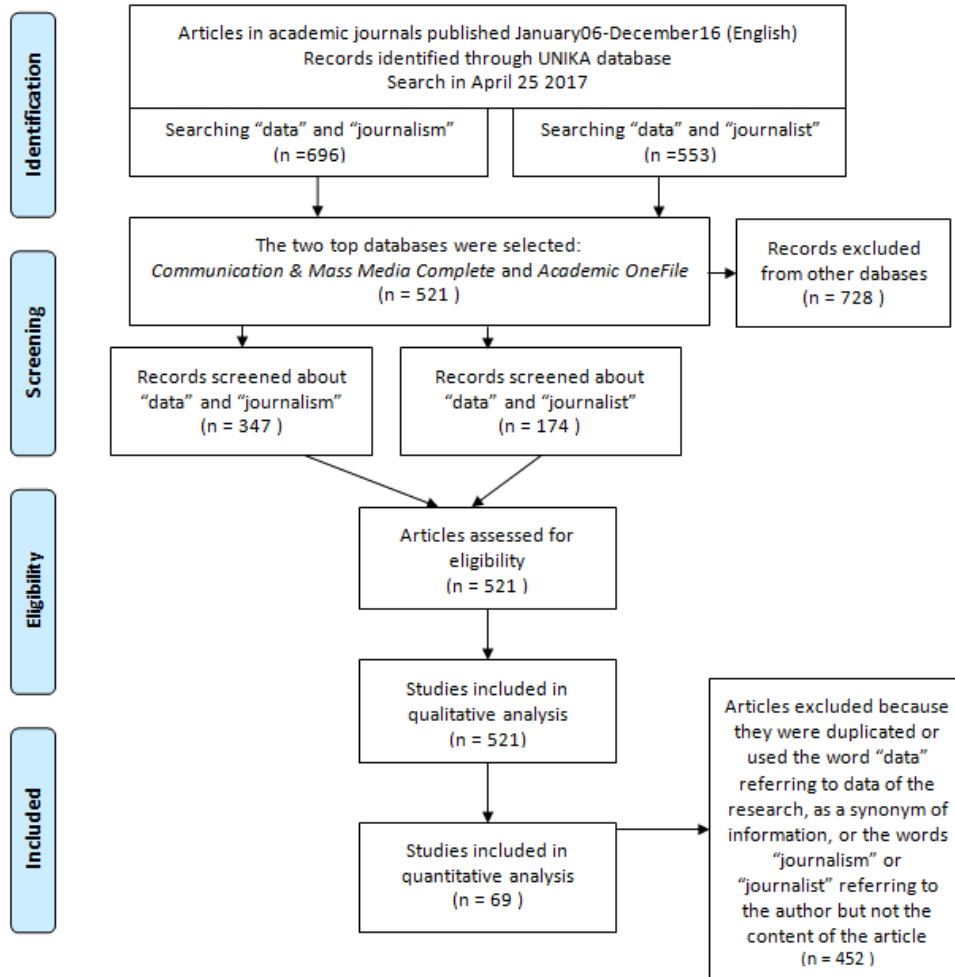
Unika permits an advanced search and the criteria considered were:

- Articles in academic journals (no news, magazines, books or professional publications).
- Published from January 2006 to December 2016.
- Language: English.

In a first search, the fourth criterion was that the abstract should include the terms "data" and "journalism". As a result, we found 696 articles. In a second search, the articles should include the terms "data" and

"journalist" in the abstract, replicating the other criteria, and we found 553 articles. The final number was 1,249 articles (see Figure 1).

Figure 1: Flow Diagram

PRISMA 2009 flow diagram. Model from (Prisma 2015) based on Moher et al. (2009) and Liberati et al. (2009).

Due to the high number of articles and the fact that only one coder would analyse the information to avoid reliability limitations (Wimmer and Dominick, 2011), only the two top databases were selected (see Table 1): *Communication & Mass Media Complete* and *Academic OneFile*.

Table 1: Databases selected, number of articles and percentage over total contributions

Databases selected	Total number of articles	Percentage over total articles found
Communication & Mass Media Complete	386	30.9%
Academic OneFile	135	10.8%
Articles found in all databases	1,249	

The databases selected included 521 articles of the 1,249, 41.7% of the total (see Table 1). The *Communication & Mass Media Complete* website (EBSCO, 2017) includes full text for more than 440 journals, abstracts for more than 670 journals and selected coverage of nearly 180 journals. *Academic OneFile* provides access to articles from over 17,000 scholarly journals and other authoritative sources such as *Nature*, *The Economist* or *The New York Times* (Gale, 2017). This last resource covers several subjects, such as physical and social sciences, technology, medicine, engineering or literature, unlike *Communication & Mass Media Complete,* which only covers communication.

After reading the titles, keywords and abstracts of the 521 articles selected from these two databases, some articles did not refer to data, journalism and journalist issues although they included the terms. Some articles used the word "data" referring to the data used for the research (a survey, for example) or as a synonym of information. Other articles used the words "journalism" or "journalist" referring to the author but not the content of the article. Furthermore, some articles were duplicates. After excluding all these cases, the final sample related to the use of data in journalism and journalistic practice comprised 69 articles (see Figure 1).

These 69 articles were included in an Excel database. For each article, the registered variables were title, authors, name of journal, year, other information about the article (issue, volume, pages), a relevant sentence from the abstract, and the keywords indicated in the search tool (from the authors or the database). In addition, we gave them a category that identified an area relating data to journalism/journalist, taking into account the ones previously highlighted and listed below:

- Source of news
- Reporting
- Services related to data
- Media business

If an article could be included in two of these areas, we associated it with both of them, including a code with the two categories.
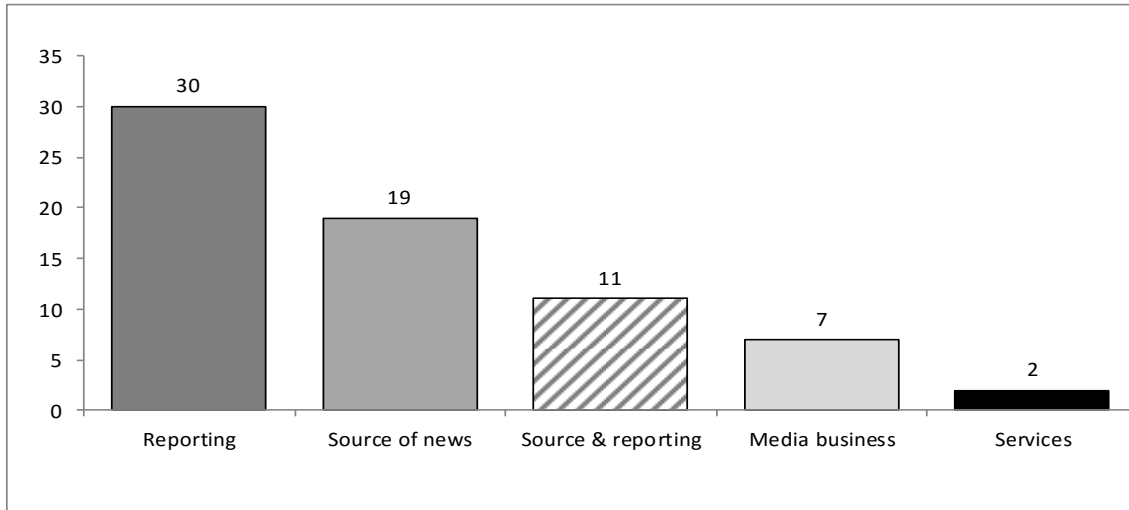
As the purpose was highlighting skills related to data management and journalistic work, we conducted a more exhaustive analysis introducing two extra thematic categories for each article. These emergent codings took into account the title, keywords and abstract of the paper. This is the most subjective process of the analysis, but it can provide more insights into the skills and knowledge required for the management of data in journalism practice.

Results

The four selected areas that link data with journalism have an impact on the academic contributions, but with a relevant prevalence of data as a source of news and as content of news (see Graph 1). The category of data as content of the news (reporting) supersedes the other, but these two categories overlap in 11 articles that talked about using data either as a source or as a part of the news. Adding this overlap, 33 articles dealt with data reporting and 27 with the use of data as a source for news, so 60 of the 69 articles
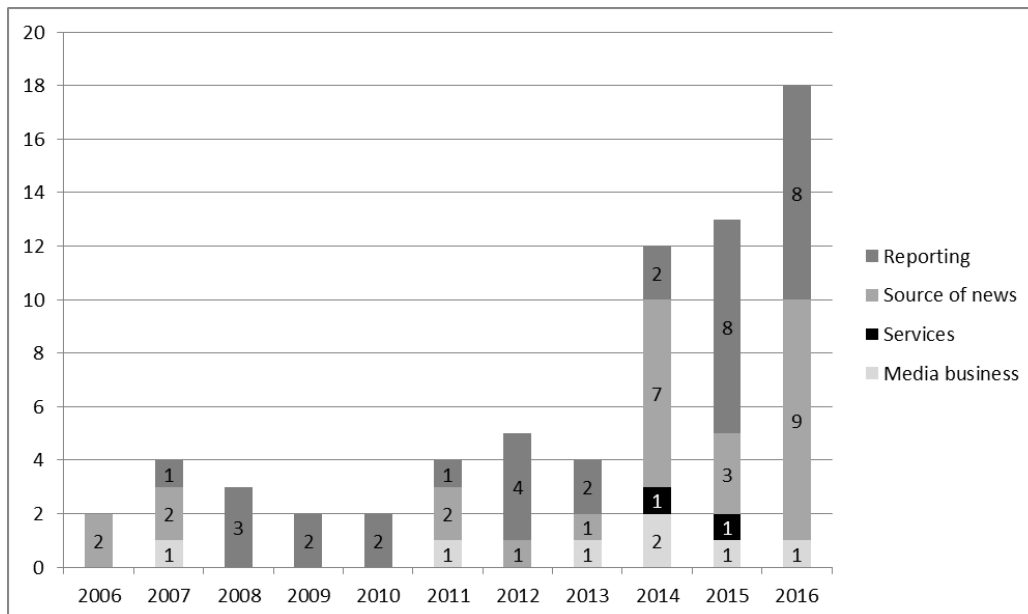
(87%) were about these two areas. Only seven cases were about data in the media business and two about services related to data (see Graph 1).

Graph 1: Number of contributions by areas that link data with journalism/journalist

Over the 69 contributions, 43 (62%) were published in the last three years, 2014 to 2016 (see Graph 2). These are the years after Wikileaks and of the development of big data. In this final three-year period, six articles talked about big data and another six about open data, the two milestones highlighted in this research. These 12 cases constitute 28% of the 43 contributions published in those years, which confirms the relevance of these two issues today.

Graph 2: Evolution of the number of contributions by areas

Besides the four general categories, we assigned 56 different ones as two extra thematic categories for each article. The most used one was "case study", including eight articles dealing with reports about a particular media brand or geographical region. Big data and open data were also relevant ones, as we have seen. Ethics, citizen journalism and social networks also had six or more contributions.

All these emergent categories were not in a previous list, and we will analyse them crossing them with the four main categories to gain insights (see Table 2).

Table 2: Emergent codes from the content analysis by areas that link data and journalism/journalist.

Main areas	Secondary categories (alphabetic order)
Data a source for news	Accuracy/Objectivity, Analytics, Automated journalism, Big data, Big data/Social networks, Case study, Citation, Citizen journalism, Crime data, Data analysis, Data journalism, Data journalist, Data reporting*, Databases, Digital data reporting , Digital journalism, Digitalization, Drones, Ethics, Mapping tools, Numeracy, Open data, Polls, Privacy, Radio, Science journalism, Social networks, Tools and analysis, Wikileaks
Data reporting	Accuracy/Objectivity, Achievements and awards, Analytics, Awards, Big data, Book review, Case study, Citizen journalism/power, Computers, Data journalist, Data security, Data-driven journalism , Difficult reporting, Digital journalism, Education, Education in data journalism, Ethics, Health, Innovation, Interactive narratives, Journalists, Laws and regulations, Mobile interface, Numeracy, Open data/Education in data journalism , Open data/Visualization Plagiarism, Poetry, Polls, Practice, Science journalism, Social networks, Source of news*, Statistics, Surveillance, Tools for images, Visualization
Services related to data	Case study, Citizen journalism, Data-driven website
Data in the media business	Audiences, Big data, Citizen journalism, Data analysis, Datafication, Digital journalism, Editor selection of news, Marketing, Markets

*Category for cases in both main groups.

In the area of data as a source of news, there are several emergent categories related to the task of "Documentation": Big data, Big data/Social networks, Crime data, Databases, Open data, Polls, Social networks.

We find also several emergent categories in the two main areas (data as a source and reporting data) related with the use of "Technology": Computers, Digitalization, Drones, Innovation or Mobile interface. Big data, Big data/Social networks. Social networks could also be in "Technology" due to its relation to it, and not only in "Documentation".

Many terms associated with the use of data as a source and data reporting are the same and can be associated directly to "Statistics" and "Visualization". Analytics, Data analysis, Numeracy, Statistics Tools and

analysis link with "Statistics". "Visualization" relates to Mapping tools, Open data/Visualization, Tools for images and Visualization

There are also many categories related to different types of specialized journalism. The most relevant specialized journalism is "Data journalism". Nine contributions included "data journalism" or "data-driven journalism" in their keywords and there are many emergent categories related with this kind of journalism: Achievements and awards, Awards, Book review, Case study, Data journalist, Data-driven journalism, Difficult reporting, Education, Education in data journalism, Journalists, Open data/Education in data journalism, Poetry or Practice. Specifically related to education there are only seven articles, four of them published in 2015-2016, which confirms the idea that "data journalism education has attracted little attention in the literature until very recently" (Hewett 2016, 120).

Other specialized journalisms cited in the articles are "Digital journalism" (Digital data reporting, Digital journalism, Computers, Interactive narratives or Mobile interface) and "Science journalism". In addition, there are case studies related to the areas of "Crime", "Health", politics or public opinion (linked with "Polls") and even with the media "Radio". There is also one contribution that cited "Automated journalism" (Splendore, 2016). This emergent data-centric practice use algorithms that convert data into narrative news texts with limited to no human intervention (Carlson 2015).

"Citizen journalism" is a special category because is related to three of the main areas: data as a source, reporting data and services related to data. This type of journalism is related to Power, Data-driven websites and also Big data and Social Networks.

If we come back to the use of data as a source and reporting data, the categories that are not mentioned in the previous paragraphs are: Accuracy/Objectivity, Citation, Data security, Ethics, Laws and regulations, Plagiarism, Privacy, Surveillance and Wikileaks. The issues related with Accuracy/Objectivity, Citation or Plagiarism link with ethics in the practice of journalism. Furthermore, journalism is an important media field concerned with privacy (McStay, 2017). Balancing the right to privacy with a free press, acting in the public interest or protecting sources are relevant issues for any journalist. Data security, Ethics, Laws and regulations, Privacy, Surveillance and Wikileaks relate to ethics in data management, especially big data. The journalists must fulfil specific ethical standards when using data as a source or reporting it.

Finally, in the area of data for business, the articles talked about Audiences, Big data, Data analysis, Datafication, Markets and Marketing, but also about Citizen journalism, Digital journalism, and the Editor selection of news. These contributions talked about tracking the audience in digital journalism, the metrics and the participation of the audience in online media. Therefore, the digitalization of media links with the management of data and big data by the media business. Journalists should learn about audience metrics and marketing, and the relevance of data to decision-making in media companies.

Limitations of the study and future research

Although the selection of the two main academic databases, *Communication & Mass Media Complete* and *Academic OneFile*, gave us a sample of 69 cases, the intervention of the researcher in choosing the body of literature to be reviewed generates a bias, as in all literature reviews (Massaro, Dumay, and Guthrie, 2016).

As we said before, the search tool used to select the contributions can vary the documents found depending on the date. Therefore, the process followed in this research is not completely replicable. Moreover, although one criterion for the advanced search was that the language of the articles should be English, some articles only used this language in the abstract.

The replication of this study is also difficult because the delimitation of the emergent categories followed an inductive process. This means that it was not an objective process, although it helped to gain better insights.

Conclusions

The development of the open data and big data is changing the journalism profession. There is more information available than ever, and journalists need new skills to manage it. The purpose of this work is to see what academics have talked about regarding data and journalism from 2006 to 2016.

Applying a structured literature review (SLR) and a content analysis with a priori and an emergent coding, 69 academic contributions about data and journalism or journalist are analysed. In the study year by year, we confirm that big data and open data are two relevant issues in journalism, but mainly after Wikileaks, in 2014.

Data can relate with journalism in different areas. In this paper, we consider four priori categories and, according to the results of the analysis, two of them are the most relevant in the academic contributions studied: data as a source for news and reporting data. The other two categories are services related to data and data for media business, and both appear in the academic papers although with less frequency.

Developing emergent categories in the content analysis, we can highlight several issues that relate data to the journalism practice. The most obvious relation is the practice of data-driven journalism. Those journalists that work with data need to have skills in statistics and visualization. The use of technology is also important as a part of documentation and for data-processing (Parasie, 2015). An adequate use of data is also required in science journalism, and for politics, public opinion, crime or health news.

Data is also key in digital journalism. The generalized use of computers and mobile phones or tablets for surfing the net implies that the media have to adapt their content to these devices, their audience and how they use them. Journalists need to manage new types of practices (such as automated journalism) and interactive narratives, both linked with the use of data in their application or the measurement of their success. The relation of data and interaction is so relevant that *the Financial Times* uses the name "Interactive news team" for the people that work in data journalism projects (Uskali and Kuutti, 2015).

The management of data is also necessary for the development of citizen journalism, in order to follow the participation of the audience and to offer them data services for their empowerment. Some data journalism projects have even asked citizens to feed the databases to improve reporting (Reisner, 1995).

The ethical issues in the practice of journalism have always been a relevant issue in journalism. Now, accuracy and objectivity are even more relevant to avoid misunderstandings when data are included in news. Besides, the management of data requires fulfilling laws and standards related to data security and privacy that journalist should know about.

Finally, the data are nowadays essential for decision-making in any company, including digital media. Journalists should learn about metrics and marketing, managing the data properly for the selection of news or their location in the media content.

In conclusion, the increase of the data available due to the open data movement and the development of big data requires new skills in journalism practice. Statistics, visualization and the use of technology have become more important, especially for the collection of data and reporting it. Furthermore, an adequate use of data is necessary for knowing the audience, developing data services and taking adequate media decisions.

References

Aguado, J.M. and Martínez, I. J (2009). Mobile Media Implicit Cultures: Towards a Characterization of Mobile Entertainment and Advertising in Digital Convergence Landscape. *Observatorio (OBS*) Journal*, 3 (1), 336-352. http://obs.obercom.pt/index.php/obs/article/view/253

Arrese, Á. (2002). *Prensa económica. De La Lloyd's List al Wsl.com*. Pamplona: Eunsa.

Ausserhofer, J. et al (2017). The datafication of data journalism scholarship: focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, April. doi: 10.1177/1464884917700667.

Beleaga, T. (2013). In a decade or less, could not all reporters requiere to be data literate? In *Data Journalism: Mapping the Future*, edited by John Mair and Richard Lance Keeble, 26–33. Suffolk, UK: Abramis Academic Publishing.

Bradshaw, P. (2010). How to be a data journalist. *The Guardian. October 1st.* https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide.

Bradshaw, P. (2014). The transparency opportunity: Holding power to account - or making power accountable? In *Transparency in Politics and the Media: Accountability and Open Government*, edited by Nigel Bowles, James T. Hamilton, and David A. L. Levy, 141–65. New York. I. B. Tauris & Co. Ltd., in association with the Reuters Institute for the Study of Journalism, University of Oxford.

Carlson, M. (2015). The robotic reporter. Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism* 3 (3). 416–31. doi: 10.1080/21670811.2014.976412.

Coddington, M. (2015). Clarifying journalism's quantitative turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* 3 (3). 331–48. doi: 10.1080/21670811.2014.976400.

Datablog (2011). The first guardian data journalism. *The Guardian.*, May 5 https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian.

Davies, T. (2010). A timeline of open government data. *Open Data Impacts. The Research Blog of @timdavies*. http://www.opendataimpacts.net/2010/06/a-timeline-of-open-government-data/.

Dawidowicz, P. (2010). *Literature reviews made easy : a quick guide to success*. Charlotte, NC: Information Age Publishing.

DeFleur, M. H. (1997). *Computer-assisted investigative reporting: Development and methodology*. Erlbaum.

Diakopoulos, N. (2012). *Cultivating the landscape of innovation in computational journalism*. http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systematicinnovation.pdf

EBSCO (2017). *Communication & Mass Media Complete*. Accessed May 25. https://www.ebscohost.com/academic/communication-mass-media-complete.

Elías-Pérez, C. (2015). *Big data y periodismo en la sociedad red*. Madrid: Síntesis.

Evens, T. and K. Van Damme (2016). Consumers' willingness to share personal data: Implications for newspapers' business models. *International Journal on Media Management* 1277 (April). 1–17. doi: 10.1080/14241277.2016.1166429.

Ferreras-Rodríguez, E. M. (2012). Nuevos perfiles profesionales: El periodista de datos. In *Actas – IV Congreso Internacional Latina de Comunicación Social – IV CILCS* – Universidad de La Laguna, Diciembre 2012, 1–19.

Forbes Insights (2014). *The promise of privacy: Respecting consumers' limits while realizing the marketing benefits of Big Data*. https://images.forbes.com/forbesinsights/StudyPDFs/turn_promise_of_privacy_report.pdf

Gale (2017). *Academic OneFile*. Accessed May 25. http://www.cengage.com/search/productOverview.do?Ntt=academic+onefile%7C%7C129663692219470930731132330063644714198&N=197&Ntk=APG%7C%7CP_EPI&Ntx=mode+matchallpartial.

Garrison, B. (1998). *Computer-Assisted Reporting*. L. Erlbaum Associates.

Gray, J., L. Bounegru and L. Chambers (2012). *The data journalism handbook: How journalists can use data to improve news*. Edited by Jonathan Gray, Liliana Bounegru, and Lucy Chambers. Sebastopol (CA): O'Reilly Media.

Hewett, J. (2013). Learning in progress: From computer-assisted reporting to data journalism, via freedom of information, open data and more. In *Data Journalism: Mapping the Future*, edited by John Mair and Richard Lance Keeble, 3–14. Suffolk, UK: Abramis Academic Publishing.

Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. *Journalism* 17 (1). 119–37. doi: 10.1177/1464884915612681.

Holtgrewe, U. and R. Werle (2001). De-commodifying software?: Open source software between business strategy. *Science Studies* 14 (2). 43–65.

Houston, B. (1996). *Computer-assisted reporting: a practical guide*. New York. St Martin's Press.

Houston, B., L. Bruzzese, and S. Weinberg (2002). *The investigative reporter's handbook: a guide to documents, databases, and techniques*. New York. Bedford/St. Martin's Press.

Howard, A. B. (2014). *The art and science of data-driven journalism.* New York: Tow center for digital journalism, Columbia Journalism School.

Hughes, N. (2013). A beginner's guide to data journalism and data mining/scraping. In *Data Journalism: Mapping the Future*, edited by John Mair and Richard Lance Keeble, 43–50. Suffolk, UK: Abramis Academic Publishing.

Hurrel, B. and J. Walton (2013). How does that affect me? Making data personally relevant for your audience. *In Data Journalism: Mapping the Future*, edited by John Mair and Richard Lance Keeble, 19–25. Suffolk, UK: Abramis Academic Publishing.

La-Rosa, L. and T. Sandoval-Martín (2016). La insuficiencia de la ley de transparencia para el ejercicio del periodismo de datos en España. *Revista Revista Latina de Comunicación Social* 71. 1208–29. doi: 10.4185/RLCS-2016-1142.

Laney, D. (2001). 3D Data management: controlling data volume, velocity, and variety | BibSonomy. *Gartner Blog*. https://www.bibsonomy.org/bibtex/263868097d6e1998de3d88fcbb7670ca6/sb3000.

Lewis, S. C. and O. Westlund (2015). Big data and journalism. Epistemology, expertise, economics, and ethics. *Digital Journalism* 3 (3): 447–66. doi: 10.1080/21670811.2014.976418.

Liberati, A. Det al (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* 6 (7). doi: 10.1371/journal.pmed.1000100.

López-Cantos, F. J. (2015). La investigación en comunicación con metodología big data. *Revista Latina de Comunicación Social* 70: 878–90. doi: 10.4185/RLCS-2015-1076.

Manovich, L. (2001). *The language of new media*. Edited by MIT Press. Cambridge (Massachusetts) . London (England): MIT Press.

Margetts, H. (2014). Data, data everywhere: Open data versus Big Data in the quest for transparency. In *Transparency in Politics and the Media: Accountability and Open Government*, edited by Nigel Bowles, James T. Hamilton, and David A. L. Levy, 167–78. New York: I. B. Tauris & Co. Ltd., in association with the Reuters Institute for the Study of Journalism, University of Oxford.

Marks, R. (2013). *The big opportunity: audience research meets big data A report for the IPA*. London. http://www.rajar.co.uk/docs/news/IPABIGOPPORTUNITY.pdf

Massaro, M., J. Dumay, and J. Guthrie (2016). On the shoulders of giants: undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal* 29 (5). Emerald Group Publishing Limited: 767–801. doi: 10.1108/AAAJ-01-2015-1939.

Mayer-Schönberger, V. and K. Cukier (2013). *Big Data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.

McStay, A. (2017). *Privacy and the media*. Los Ángeles (etc.): Sage.

Meyer, P. (2001). *The new precision journalism*. Lanham, MD: Rowman and Littlefield.

Moher, D. et al (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine* 6 (7). doi: 10.1371/journal.pmed.1000097.

Napoli, P. M. (2016). Special issue introduction: Big data and media management. *International Journal on Media Management* 18 (1). 1–7. doi: 10.1080/14241277.2016.1185888.

Open Government Partnership (2016). *Open Government Partnership. Open by Default, Policy by the People, Accountability for Results*. Washington: UGD.

Papí-Gálvez, N. (2014). Los medios online y la ¿crisis? De la planificación de medios publicitarios. *adComunica* (7). 29–48. doi:10.6035/166.

Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of "big data". *Digital Journalism* 3 (3). 364–80.

Parasie, S. and Eric Dagiral (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society* 15 (6). 853–71. doi: 10.1177/1461444812463345.

Prisma (2015). *PRISMA. Transparent reporting of systematic reviews and Meta-analyses*. Ottawa Hospital Research Institute and University of Oxford.

Reisner, N. H. (1995). On the beat: Computer-assisted reporting isn't just for projects anymore. American Journalism Review, 17(2), 44–47.

Royal, C. and D. Blasingame (2015). Data journalism: An explication. *International Symposium in Online Journalism* 5 (1). 24–46.

Silva, A., X. López, O. Westlund. and N.G. Ulloa (2016). Iniciativas en la comunicación móvil: Mapeo de los canales, el contenido, la conversación y el comercio en Europa y los EE.UU. *Observatorio (OBS*) Journal*, 10(1), 201-217. http://obs.obercom.pt/index.php/obs/article/view/889

Splendore, S. (2016). Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass* 10 (5). 343–52. doi: 10.1111/soc4.12366.

Teruel, L. and E. Blanco (2015). La incipiente apuesta por el periodismo de datos en el análisis político y electoral español: Las Elecciones catalanas de 2015. In *La Pantalla Insomne*, 1107–28. Universidad de la Laguna. doi: 10.4185/cac90.

Tranfield, D., D. Denyer, and P. Smart (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14: 207–22.

Uskali, T. and H. Kuutti (2015). Models and streams of data journalism. *The Journal of Media Innovations* 2 (1). 77–88.

Westlund, O. (2013). Mobile news. A review and model of journalism in an age of mobile media. *Digital Journalism* 1 (1). Routledge: 6–26. doi: 10.1080/21670811.2012.740273.

Wimmer, R. D. and J. R. Dominick (2011). *Mass media research: an introduction*. Cengage- Wadsworth.